

Список публикаций:

- [1] Борисова З.У. Халькогенидные полупроводниковые стекла. – Л.: Изд-во Ленингр. ун-та. 1983. С. 344.
- [2] Борисова З.У. Химия стеклообразных полупроводников. – Изд-во Ленинградского ун-та. 1972. С. 248.
- [3] Tanaka K., Shimakawa K. *Amorph. Chalc. Semicond. And Related Mat.* NY:Springer. 2011.
- [4] Мельникова Н. В., Курочка К. В., Хейфец О. Л., Кадырова Н. И., Волкова Я. Ю. Атомная структура и ионная проводимость стеклообразных материалов на основе сульфида серебра // ИЗВЕСТИЯ РАН. СЕРИЯ ФИЗИЧЕСКАЯ. 2015. Т.79. № 6. С. 790–794.
- [5] Kurochka K.V., Melnikova N.V. Investigation of electrical properties of glassy $\text{AgGe}_{1-x}\text{As}_x\text{I}-x(\text{S}+\text{CNT})_3$ ($x = 0.4; 0.5; 0.6$) at temperature range from 10 to 300 K // *Solid State Ionics*. 2017. V.300. P. 53-59.

Новая элементная база для нейроморфных вычислений

Исламов Дамир Ревинирович

Институт физики полупроводников им. А. В. Ржанова СО РАН

Новосибирский национальный исследовательский государственный университет

damir@isp.nsc.ru

Улучшения литографии в технологическом процессе производства полупроводниковых микросхем позволяли увеличивать число транзисторов экспоненциальными темпами. Впервые это заметил директор по НИОКР в «Fairchild Semiconductor» Гордон Мур. В 1965 году (через шесть лет после изобретения интегральной схемы) он написал внутренний доклад «Будущее интегральной электроники» с графиком, соединяющим 5 точек и связывающим число компонентов интегральных схем и их минимальную цену для периода 1959-1964 гг., и предсказанием развития на следующие 10 лет. Последнее основывалось на том, что число компонентов на кристалле будет продолжаться удваиваться каждый год. 19 апреля 1965 г. отредактированная версия появилась в виде статьи в журнале «Electronics». [1]. Это наблюдение получило название – закон Мура.

В 1975 году на ежегодной встрече Международной Организации Инженеров-Электронщиков (IEEE) Гордон Мур, будучи президентом и исполнительным директором «Intel» указал, что увеличение диаметра пластин, успехи в технологических процессах и «поумнение схем и устройств» позволило продолжиться прогнозу. Впрочем, Мур скорректировал свою закономерность до удвоения степени интеграции каждые 2 года. Предсказание оказалось самоподдерживающимся: теперь Гордон Мур уже не наносит на свой график очередные достижения, зато многочисленные фирмы (и, конечно, сама «Intel») до сих пор стараются идти в ногу с прогрессией. В 1995 году, ещё раз публично проверяя своё предсказание, Мур сделал вывод, что оно «не скоро остановится».

Рисунок 1 иллюстрирует экспоненциальный рост производительности выпускаемых логических микросхем, нормированных на их стоимость. Если прост продолжится, после 2020 года производительность микропроцессоров должна достичь уровня мозга живого объекта, а к 2045 году превзойти ментальную производительность всего человечества на Земле. В связи с этим наиболее перспективным путём развития электроники видится внедрение новых физических принципов, например, нейроморфные вычисления.

Самый производительный суперкомпьютер в мире (по состоянию на июнь 2016 года) со скоростью вычислений 93 петафлопс Sunway TaihuLight в Национальном суперкомпьютерном центре города Уси (провинция Цзянсу, КНР) для работы требует 15,371 МВт электроэнергии (вместе с системой охлаждения – 28 МВт). Для сравнения человеческий мозг содержит 10^{11} нейронов ($\sim 10^6$ нейронов/см²) и 10^{15} синапсов ($\sim 10^{10}$ синапсов/см²), в пике активности потребляет 20 Вт. Характерная частота работы мозга 10 Гц. За счёт параллелизма взаимодействия синапсов мозг совершает $10^{15} \times 10 \text{ Гц} = 10^{16}$ операций в секунду. Энергетический выигрыш при работе биологического объекта – мозга – по сравнению с самым мощным суперкомпьютером $\sim 10^5$. Мозг может за доли секунды быстро выполнять задачи, на которые компьютеру потребуются дни или даже месяцы. Среди таких задач можно отметить распознавание графических или звуковых образов.

Секрет такой разницы заключается в архитектуре. Современные вычислительные машины построены по архитектуре фон Неймана, принципа совместного хранения команд и данных в памяти вычислительной системы. Сами же вычисления проводятся в вычислительном модуле. Данные и инструкции из памяти передаются в вычислитель через шину, быстродействие которой и ограничивает вычислительные мощности компьютера. Архитектура мозга драматически отличается от архитектуры фон Неймана. Память и вычислитель здесь являются одними и теми же объектами – синапсами, местом соединения нейронов, составляющих мозг. Для передачи данных и инструкций в синаптический вычислитель какая-либо шина не нужна [2].

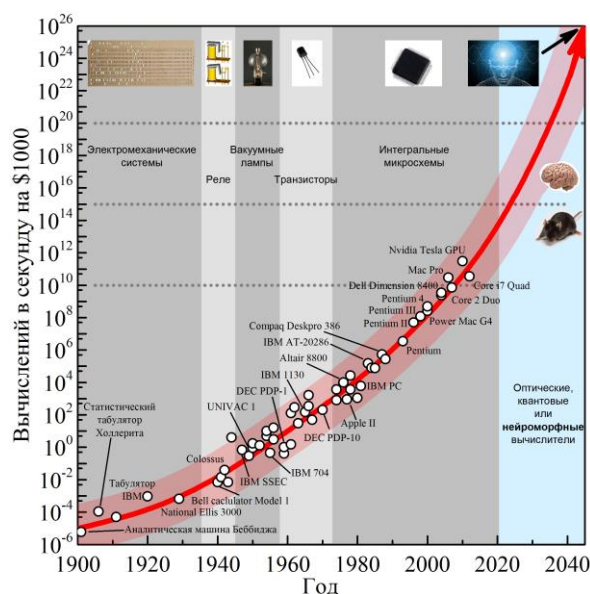


рис.1. Количество выполняемых логических операций в секунду, нормированное на стоимость микросхем, выпущенных в разные годы. Красная линия иллюстрирует закон Мура. Чёрная стрелка указывает на предел роста, соответствующий вычислительным мощностям мозга условного коллективного разума всего человечества

Несмотря на то, что проблемой искусственного интеллекта учёные занимаются достаточно давно, большинство современные реализации искусственных нейронов основаны на программном коде, выполняющимся на обычных фон Неймановских машинах. Для наибольшей эффективности нейронной сети естественно предположить, что реализовывать её надо не на программном уровне, а в виде специальной вычислительной машины. В 2011 году подразделением «IBM Research» был представлен научно-технической аудитории первый кристалл с одним нейросинаптическим ядром, 256 программируемых нейронов и 262144 программируемых синапсов. 7 августа 2014 года «IBM Research» продемонстрировали новый когнитивный процессор «TrueNorth» с архитектурой «SyNAPSE». Нейропроцессор содержал 4096 нейросинаптических ядер (в двухмерном массиве размером 64×64), включавших один миллион программируемых нейронов и 256 миллионов программируемых синапсов. Энергопотребление составило 63 мВт. Каждый нейрон каждого из ядер может посылать сигнал к каждому нейрону из другого ядра. После этого сигнал по программируемым синапсам разбегается и по прочим нейронам в этом ядре, в соответствии с установленными или установившимися там весовыми коэффициентами. В качестве примера работы когнитивного нейропроцессора сотрудники «IBM» продемонстрировали классификацию объектов (людей, велосипедистов, машины, грузовики и автобусы) в видеопотоке 400×240 пикселей с частотой 30 кадров в секунду [3]. Процессор TrueNorth был изготовлен по планарной полупроводниковой технологии с проектной нормой 28 нм на заводе «Samsung», содержит 5,4 миллиарда транзисторов, задействованных в эмуляции нейронов и синапсов, а также в хранении данных в ячейках СОЗУ памяти суммарным объёмом 400 Мбит. Синапсы имеют веса, выражающиеся как 9-битные целые, и 4-битную временную задержку. В роли нейрона выступает стандартный сумматор на основе операционного усилителя с отрицательной обратной связью. Несмотря на все достоинства TrueNorth, он обладает одним ключевым недостатком — энергозависимостью. При выключении питания все инструкции и данные теряются безвозвратно.

Чтобы нейрокомпьютер избавить от энергозависимости, блоки для обработки и хранения данных необходимо проектировать на новой элементной базе. И такие элементы были найдены. Сначала теоретически, а после и экспериментально. На текущий момент синаптическая пластичность была продемонстрирована ячейками на основе мемристора (пассивного элемента, изменяющего свой сопротивление в зависимости от значения и направления протекающего тока) и сегнетоэлектрического элемента памяти. Принцип действия обоих этих приборов основан на перестройке атомной структуры. В мемристоре при переключении из высокоомного состояния в низкоомное (и обратно) дефекты формируются в определённой пространственной конфигурации. В сегнетоэлектрическом элементе атомы перестраиваются с формированием сонаправленных электрических диполей.

Настоящая работа посвящена проблемам и задачам, стоящими перед разработчиками перспективной элементной базы для нейроморфных вычислений.

Список публикаций:

[1] G. E. Moore. Cramming more components onto integrated circuits // *Electronics*, 38(8), 114–117 (1965).

[2] M. Versace, B. Chandler. The brain of a new machine // *IEEE Spectrum*, 47(12), 30–37 (2010).

[3] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface // *Science*, 345(6197), 668–673 (2014).